

A Theory of Constant $E\tau^2$ CMOS Circuits

Karl Papadantonakis

**Computer Science Department
California Institute of Technology**

caltechCSTR/2001.004

A Theory of Constant $E\tau^2$ CMOS Circuits

Karl Papadantonakis

9/19/2001

Introduction

In a digital CMOS circuit, used digitally, the dynamic energy consumption of a single pulldown stage has the form CV^2 , while the current i is kV^2 , and the delay τ is CV/i , where C is an output capacitance, V is the supply voltage, and k is a transconductance parameter.

Recently Martin proposed using $E\tau^2$ as a voltage-independent circuit performance metric [AM]; indeed we have:

$$E\tau^2 = (CV^2) \cdot \left(\frac{CV}{i}\right)^2 = (CV^2) \cdot \left(\frac{CV}{kV^2}\right)^2 = \frac{C^3}{k^2}, \quad (1)$$

which is independent of voltage. In particular, the speed of the pulldown network can be changed without changing $E\tau^2$.

Previously, this argument about single pulldown stages was applied to arbitrary digital circuits by assuming that the total delay of any such circuit was a sum of single pulldown and pullup stage delays, neglecting the time variation of the input to each stage. In this paper we show how to generalize the $E\tau^2$ metric to arbitrary networks of CMOS transistors using only a weaker assumption: that every node has a capacitance.

Acknowledgements

Endless help from and consultation with my advisor, Alain Martin. Additional helpful comments from Mika Nyström and Martin Rem.

MOS Equations

To first order, any NMOS transistor is described by $i_G = 0$ and the following equations:

$$i_D = -i_S = \begin{cases} i_l = k \left[(v_{GS} - v_T)v_{DS} - \frac{v_{DS}^2}{2} \right], & \text{if } v_{DS} < v_{GS} - v_T \text{ and } v_{GS} > v_T; \\ i_{fs} = \frac{k}{2}(v_{GS} - v_T)^2, & \text{if } v_{DS} > v_{GS} - v_T > 0; \\ i_{rs} = -\frac{k}{2}(v_{GD} - v_T)^2, & \text{if } v_{SD} > v_{GD} - v_T > 0; \\ 0, & \text{if } v_{GS} \leq v_T \text{ and } v_{GD} \leq v_T. \end{cases} \quad (2)$$

The inequalities are reversed for PMOS transistors. v_T is the threshold voltage, v_{GS} is the gate-to-source voltage, v_{GD} is the gate-to-drain voltage, and $v_{DS} \stackrel{\text{def}}{=} v_{GS} - v_{GD}$. k and v_T are fixed at fabrication, and are both positive for NMOS transistors, and both negative for PMOS transistors.

Equation 2 is the standard form of the MOS equations [NM]. However, these equations are hard to read: First, it is not even clear that they are consistent. Secondly, even though there is no physical distinction between D and S in a fabricated MOS transistor, it is not evident that the equations are unchanged if D and S are interchanged. The equations are much more clearly and symmetrically written in terms of v_{GS} and v_{GD} alone. Rewriting the first case, we have

$$\begin{aligned} k \left[(v_{GS} - v_T)v_{DS} - \frac{v_{DS}^2}{2} \right] &= -\frac{k}{2} [v_{DS}^2 - 2(v_{GS} - v_T)v_{DS} + (v_{GS} - v_T)^2 - (v_{GS} - v_T)^2] \\ &= -\frac{k}{2} [(v_{DS} - v_{GS} + v_T)^2 - (v_{GS} - v_T)^2] \\ &= \frac{k}{2}(v_{GS} - v_T)^2 - \frac{k}{2}(v_{GD} - v_T)^2, \end{aligned} \quad (3)$$

so that Equation 2 becomes

$$i_D = -i_S = \begin{cases} i_l = \frac{k}{2}(v_{GS} - v_T)^2 - \frac{k}{2}(v_{GD} - v_T)^2, & \text{if } v_{GS} > v_T \text{ and } v_{GD} > v_T; \\ i_{fs} = \frac{k}{2}(v_{GS} - v_T)^2, & \text{if } v_{GS} > v_T \text{ and } v_{GD} \leq v_T; \\ i_{rs} = -\frac{k}{2}(v_{GD} - v_T)^2, & \text{if } v_{GS} \leq v_T \text{ and } v_{GD} > v_T; \\ 0, & \text{if } v_{GS} \leq v_T \text{ and } v_{GD} \leq v_T. \end{cases}, \quad (4)$$

or, more simply:

$$i_D = -i_S = i_f - i_r, \quad (5)$$

where

$$\begin{aligned} i_f &\stackrel{\text{def}}{=} \begin{cases} \frac{k}{2}(v_{GS} - v_T)^2, & \text{if } v_{GS} > v_T; \\ 0, & \text{otherwise.} \end{cases} \\ i_r &\stackrel{\text{def}}{=} \begin{cases} \frac{k}{2}(v_{GD} - v_T)^2, & \text{if } v_{GD} > v_T; \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \quad (6)$$

As before, reverse the inequalities for PMOS transistors. Henceforth we shall use this form of the MOS equations. Notice how Equations 5-6 remain unchanged if D and S are interchanged.

The Square Law for MOS transistors

The drain current i_D described in Equations 5-6 is a *transconductance function* of three *transistor* voltages: v_{GS} , v_{GD} , and v_T . In vector notation,

$$i_D = f_D(\mathbf{v}_{\text{MOS}}), \text{ where } \mathbf{v}_{\text{MOS}} \stackrel{\text{def}}{=} \begin{bmatrix} v_{GS} \\ v_{GD} \\ v_T \end{bmatrix}. \quad (7)$$

Square Law. For any transistor voltage vector \mathbf{v}_{MOS} and positive scalar s , the function f_D representing the drain current $i_D = f_D(\mathbf{v}_{\text{MOS}})$ of any MOS transistor has the property that

$$f_D(s\mathbf{v}_{\text{MOS}}) = s^2 f_D(\mathbf{v}_{\text{MOS}}). \quad (8)$$

Proof: The following lemma allows us to consider $i_f(\mathbf{v}_{\text{MOS}})$ and $i_r(\mathbf{v}_{\text{MOS}})$ separately:

Square Law Lemma. For arbitrary functions \mathbf{f} and \mathbf{g} satisfying the square laws $\mathbf{f}(s\mathbf{v}) = s^2\mathbf{f}(\mathbf{v})$ and $\mathbf{g}(s\mathbf{v}) = s^2\mathbf{g}(\mathbf{v})$, the functions $\mathbf{f} \pm \mathbf{g}$ also satisfy the square law, i.e. $(\mathbf{f} \pm \mathbf{g})(s\mathbf{v}) = s^2(\mathbf{f} \pm \mathbf{g})(\mathbf{v})$.

Proof:

$$(\mathbf{f} \pm \mathbf{g})(s\mathbf{v}) = \mathbf{f}(s\mathbf{v}) \pm \mathbf{g}(s\mathbf{v}) = s^2\mathbf{f}(\mathbf{v}) \pm s^2\mathbf{g}(\mathbf{v}) = s^2(\mathbf{f} \pm \mathbf{g})(\mathbf{v}). \quad \square \quad (9)$$

Continuing with the proof of the square law,

$$\begin{aligned} i_f(s\mathbf{v}_{\text{MOS}}) &= \begin{cases} \frac{k}{2}(sv_{GS} - sv_T)^2, & \text{if } sv_{GS} > sv_T; \\ 0, & \text{otherwise.} \end{cases} \\ &= s^2 \begin{cases} \frac{k}{2}(v_{GS} - v_T)^2, & \text{if } v_{GS} > v_T; \\ 0, & \text{otherwise.} \end{cases} \\ &= s^2 i_f(\mathbf{v}_{\text{MOS}}), \end{aligned} \quad (10)$$

using the assumption that s is positive. Interchanging S and D , the same holds for i_r . Applying the lemma to Equation 5, the result holds for i_D . Also one can reverse the inequalities, so that the result applies for PMOS transistors as well as NMOS. \square

Smooth Circuits

Eventually we will generalize the square law (Equation 8) to arbitrary networks of MOS transistors; we will show that the operating point of such a circuit is a vector of *circuit voltages* \mathbf{v} and *circuit currents* $\mathbf{i} = \mathbf{f}(\mathbf{v})$ such that \mathbf{f} satisfies the square law. However, first we must consider three problems with plain networks of MOS transistors:

1. Without any capacitance, there is no notion of time. The circuit does not compute anything more than a steady-state operating point.
2. Without any capacitance, once some of the voltages in the circuit are specified the other voltages are determined. Thus it is not obvious that one can scale all voltages in the circuit together.
3. Unfortunately, in general those other voltages are not always determined uniquely. For example, the voltage transfer curve of an inverter usually has a vertical segment.

There are several ways of getting around these problems. Our first solution will be to use what we call *smooth circuits*: closed CMOS circuits with capacitors to ground added. For simplicity suppose that “ground” is not a node in the original circuit; replace any references to “ground” in the original circuit by a 0V power supply.

We start with a network consisting solely of MOS transistors. We add to each node j in the circuit a capacitance C_j between j and ground. In practice the C_j attached to dynamic nodes may be very small; nonetheless they always exist. In addition to solving the above problems, these capacitors are a mathematical convenience: they allow power supplies and input values to be simply encoded as large capacitors with the appropriate initial values. Thus our circuits are *closed*, i.e. every circuit element is fully described; there are no “external” elements.

Now for each j let v_j denote the voltage across C_j , i.e. the absolute voltage of j . And let i_j denote the current entering capacitor C_j . For each j , we have:

$$\frac{i_j}{C_j} = \frac{dv_j}{dt}, \quad (11)$$

by the constitutive relation for a capacitor. We will use the following vector notation:

$$\mathbf{i} \stackrel{\text{def}}{=} \begin{bmatrix} 0 \\ i_1 \\ i_2 \\ \vdots \\ i_n \end{bmatrix}, \quad \mathbf{v} \stackrel{\text{def}}{=} \begin{bmatrix} v_T \\ v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}, \quad \text{and} \quad \overrightarrow{\left(\frac{1}{C}\right)} \stackrel{\text{def}}{=} \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{C_1} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{C_2} & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{C_n} \end{bmatrix}, \quad (12)$$

where n is the number of nodes in the network. Each i_j/C_j can be expressed as a function $f_j(\mathbf{v})$ of all other *circuit voltages* \mathbf{v} . As denoted above, the i_j comprise the *circuit currents* \mathbf{i} . The f_j comprise a function \mathbf{f} , where

$$\mathbf{f}(\mathbf{v}) \stackrel{\text{def}}{=} \overrightarrow{\left(\frac{1}{C}\right)} \mathbf{i} = \frac{d\mathbf{v}}{dt}. \quad (13)$$

Notice that the circuit equations always refer to $1/C_j$, never to C_j . This allows power supplies to be represented as infinite capacitances. Also the constancy of v_T is essentially achieved by an infinite capacitance, as denoted by the zero in the top left entry of $\overrightarrow{\left(\frac{1}{C}\right)}$.

Definition. A *strict global transconductance function* is a function of circuit voltages whose value is a vector of circuit currents, in which for every node in the circuit there is one scalar component equalling the total current leaving all MOS transistors connected to that node.

The full generality of the above definition will become useful later when we discuss adding more capacitors to the circuit.

According to the following weaker definition, \mathbf{f} is a *global transconductance function*:

Definition. A *global transconductance function* is any function expressible as a diagonal matrix times a *strict global transconductance function*.

The Square Law for Global Transconductance Functions

Next we extend the square law (Equation 8) to global transconductance functions.

Square Law for Strict Global Transconductance Functions. *Any strict global transconductance function \mathbf{i} satisfies the following property: for any vector \mathbf{v} assigning a voltage to every node in the circuit, and any positive scalar s , the function \mathbf{i} obeys*

$$\mathbf{i}(s\mathbf{v}) = s^2\mathbf{i}(\mathbf{v}). \quad (14)$$

Proof: Since \mathbf{i} is a strict global transconductance function, each $i_j(\mathbf{v})$ is the total current leaving all MOS transistors connected to that node. By Kirchoff's current law, i_j is the sum and difference of a finite set of drain currents i_D . Each i_D depends on some voltages \mathbf{v}_{MOS} whose indices are a subset of the indices of \mathbf{v} . Each $i_D = f_D(\mathbf{v}_{\text{MOS}})$ satisfies a square law, according to Equation 8. Consequently each $i_D = f_D(\mathbf{v})$ thought of as a function of all voltages in the circuit satisfies the square law:

$$f_D(s\mathbf{v}) = f_D(s\mathbf{v}_{\text{MOS}}) = s^2 f_D(\mathbf{v}_{\text{MOS}}) = s^2 f_D(\mathbf{v}). \quad (15)$$

The square law lemma (Equation 9) states that any sum and difference of these f_D also satisfies the square law, hence each i_j satisfies the square law. \square

Square Law for Global Transconductance Functions. *Any global transconductance function obeys*

$$\mathbf{f}(s\mathbf{v}) = s^2\mathbf{f}(\mathbf{v}). \quad (16)$$

Proof: Each $f_j = \alpha_j i_j$, where \mathbf{i} is a strict global transconductance function. For any \mathbf{v} ,

$$f_j(s\mathbf{v}) = \alpha_j i_j(s\mathbf{v}) = \alpha_j s^2 i_j(\mathbf{v}) = s^2 f_j(\mathbf{v}). \quad \square \quad (17)$$

In particular, the function \mathbf{f} defined in Equation 13 describing smooth circuits is a global transconductance function, so this form of the square law applies to smooth circuits.

Uniqueness of Solutions to Smooth Circuits

Consider an arbitrary CMOS circuit. If we assume that the circuit is smooth, i.e. the circuit is closed and a capacitor to ground is added to every node of the circuit, then the *solution of the circuit* is the vector $\mathbf{v}(t)$ of signals that evolve together according to Equation 13. In other words, given a global transconductance function \mathbf{f} describing the circuit, a solution of the circuit $\mathbf{v}(t)$ satisfies

$$\frac{d\mathbf{v}(t)}{dt} = \mathbf{f}(\mathbf{v}(t)). \quad (18)$$

Now we justify the previously presented claim that smooth circuits have unique solutions, i.e. for each \mathbf{v}_0 there is a unique $\mathbf{v}(t)$ defined for $t \geq 0$ with $\mathbf{v}(0) = \mathbf{v}_0$. By the existence and uniqueness theorems for ordinary differential equations [HW] it suffices for \mathbf{f} to satisfy a Lipschitz condition

$$\|\mathbf{f}(\mathbf{v}_1, t) - \mathbf{f}(\mathbf{v}_2, t)\| \leq K \|\mathbf{v}_1 - \mathbf{v}_2\|, \text{ for all } t, \mathbf{v}_1, \text{ and } \mathbf{v}_2, \quad (19)$$

for some Lipschitz constant K . Since our \mathbf{f} is a polynomial function of \mathbf{v} that does not depend on t , it suffices to show that any solution $\mathbf{v}(t)$ satisfies

$$\|\mathbf{v}(t)\| \leq v_{\text{MAX}}, \text{ for all } t \geq 0, \quad (20)$$

for some constant v_{MAX} . This constant exists because smooth circuits are *globally weakly restoring*: extreme voltages can only be attracted towards other voltages, if at all. Pick a time t . At that time, of all nodes, consider a node j whose voltage is of maximum value. Consider all transistors whose sources or drains connect to this node. Since MOS transistors never supply power, these transistors can only have current from the more positive terminal to the less positive terminal; thus this current, if any, can only lower the voltage at j . By a similar argument, the most negative circuit voltage can only increase. Thus

$$v_{\text{MAX}} \stackrel{\text{def}}{=} \max_j |v_j(0)| \quad (21)$$

satisfies Equation 20, even in the presence of infinite capacitances.

The Voltage Scaling Law for Smooth Circuits

Consider any smooth circuit. The following law describes how global voltage scaling affects the solutions to Equation 18:

Voltage Scaling Law for Smooth Circuits. *If $\mathbf{v}(t)$ is a solution to Equation 18, and s is any positive scalar, then $\mathbf{u}(t) \stackrel{\text{def}}{=} s\mathbf{v}(st)$ is also a solution to Equation 18.*

Proof: Suppose $\mathbf{v}(t)$ is a solution to Equation 18. Using this fact along with the definition of \mathbf{u} , the chain rule, and the square law for global transconductance functions (Equation 16), we have:

$$\frac{d\mathbf{u}(t)}{dt} = \frac{d(s\mathbf{v}(st))}{dt} = s \frac{d\mathbf{v}(st)}{d(st)} \frac{d(st)}{dt} = s^2 \mathbf{f}(\mathbf{v}(st)) = \mathbf{f}(s\mathbf{v}(st)) = \mathbf{f}(\mathbf{u}(t)). \quad \square \quad (22)$$

Voltage Scaling Law for Smooth Circuits with Additional Capacitors

The smooth circuits we have considered so far are *ordinary smooth circuits*: only capacitors with one terminal grounded are allowed. Often engineers are interested in other kinds of capacitors. For example, significant capacitance may be present between the gate of a transistor and the other two terminals, either or both of which may not be grounded.

In order to consider arbitrary additional capacitors, we must redefine the i_j so that they comprise a global transconductance function. The previous definition (Equation 11) no longer leads to a global transconductance function when additional capacitors are allowed. In fact we define i_j as the value of the *strict* transconductance function: the net current from the transistors connected to j to the capacitors connected to j :

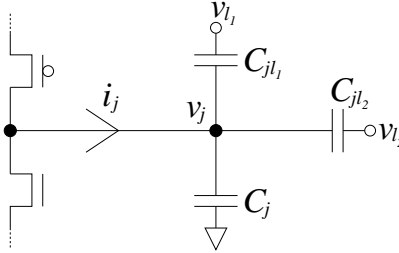


Figure 23. Example illustrating the definition of i_j in the presence of multiple capacitors.

Let C_{jl} denote the added capacitance between nodes j and l , or zero if no capacitor exists directly between these nodes. Consider the strict global transconductance function, \mathbf{f} . By definition, we have componentwise:

$$\begin{aligned} f_j(\mathbf{v}) = i_j &= C_j \frac{dv_j}{dt} + \sum_l C_{jl} \frac{d(v_j - v_l)}{dt} \\ &= \left(C_j + \sum_m C_{jm} \right) \frac{dv_j}{dt} - \sum_l C_{jl} \frac{dv_l}{dt} , \end{aligned} \quad (24)$$

or in vector notation:

$$[C] \frac{d\mathbf{v}(t)}{dt} = \mathbf{f}(\mathbf{v}(t)), \quad (25)$$

where

$$[C]_{jl} \stackrel{\text{def}}{=} \left(C_j + \sum_m C_{jm} \right) \delta_{jl} - C_{jl}. \quad (26)$$

As with the circuit equation for ordinary smooth circuits (Equation 18), we refer to any solution $\mathbf{v}(t)$ of Equation 25 as a *solution of the circuit*. And in much the same spirit as the voltage scaling law for ordinary smooth circuits, we have:

Voltage Scaling Law for Smooth Circuits with Additional Capacitors. *If $\mathbf{v}(t)$ is a solution to Equation 25, and s is any positive scalar, then $\mathbf{u}(t) \stackrel{\text{def}}{=} s\mathbf{v}(st)$ is also a solution to Equation 25.*

Proof: Suppose $\mathbf{v}(t)$ is a solution to Equation 25. Using this fact along with the definition of \mathbf{u} , the chain rule, and the square law for global transconductance functions (Equation 16), we have:

$$[C] \frac{d\mathbf{u}(t)}{dt} = [C] \frac{d(s\mathbf{v}(st))}{dt} = s \frac{d([C]\mathbf{v}(st))}{d(st)} \frac{d(st)}{dt} = s^2 \mathbf{f}(\mathbf{v}(st)) = \mathbf{f}(s\mathbf{v}(st)) = \mathbf{f}(\mathbf{u}(t)). \quad \square \quad (27)$$

For ordinary smooth circuits, the scaled solutions were unique, because all solutions are unique. However, when additional capacitors are added, it is not clear that solutions are unique, because the circuits are no longer globally weakly restoring; in fact, for these circuits, Equation 21 is emphatically false: “charge pump” circuits can be specially designed to take advantage of the extra capacitors in order to produce voltages that eventually exceed all supply voltages by orders of magnitude.

Instead of a proof of uniqueness, we will use a corollary demonstrating that once a circuit designer provides a “reference solution” that is known to be unique, the scaled solutions are also unique:

Voltage Scaling Law Corollary. *If $\mathbf{v}(t)$ is the unique solution to Equation 25 with $\mathbf{v}(0) = \mathbf{v}_0$, and s is any positive scalar, then $\mathbf{u}(t) \stackrel{\text{def}}{=} s\mathbf{v}(st)$ is the unique solution to Equation 25 with $\mathbf{u}(0) = s\mathbf{v}_0$.*

Proof: By the scaling law, $\mathbf{u}(t)$ is a solution to Equation 25, and $\mathbf{u}(0) = s\mathbf{v}_0$. Now suppose that \mathbf{w} is also a solution, and $\mathbf{w}(0) = \mathbf{u}(0) = s\mathbf{v}_0$. By the scaling law again, $\mathbf{w}(t)/s$ is a solution. Since $\mathbf{w}(0)/s = \mathbf{v}_0$ and \mathbf{v} is the unique solution with this property, we have $\mathbf{w}(t)/s = \mathbf{v}(t)$ for all t . Hence $\mathbf{w} = \mathbf{u}$. \square

Note that this corollary could also have been applied to the equation of ordinary smooth circuits (Equation 18), had it been infeasible to prove that ordinary smooth circuits have unique solutions.

Constant $E\tau^2$ as a consequence of the Voltage Scaling Law

We finally demonstrate that any ordinary smooth circuit with any initial condition \mathbf{v}_0 can perform the computation corresponding to \mathbf{v} at arbitrary speed, while $E_{\mathbf{v}}\tau_{\mathbf{v}}^2$ remains constant, where $\tau_{\mathbf{v}}$ is latency and $E_{\mathbf{v}}$ is energy.

Definition 28. *For an arbitrary smooth circuit, designate an output node out . Consider a solution $\mathbf{v}(t)$ of the circuit. If $v_{out}(t) > 0$ for some $t \geq 0$, then define the computational latency $\tau_{\mathbf{v}}$ as the greatest lower bound on all such t . If $\tau_{\mathbf{v}}$ exists, then we say that the circuit halts on initial condition $\mathbf{v}(0)$.*

The halting property can be attributed to $\mathbf{v}(0)$ alone because $\mathbf{v}(t)$ is unique once $\mathbf{v}(0)$ is specified. In general one may be interested in more information about the circuit solution than simply the time of halting. Therefore we have:

Assumption 29. *If there exists a scalar α and a positive scalar s such that two solutions $\mathbf{u}(t)$ and $\mathbf{v}(t)$ of a circuit satisfy $\mathbf{u}(t) = \alpha\mathbf{v}(st)$ for all $t \geq 0$, then we can say that $\mathbf{u}(t)$ and $\mathbf{v}(t)$ perform the same computation.*

Until now we have said nothing about computation. Assumption 29 is a reasonable property for a computational model of CMOS circuits to have. For example, in the QDI design style [AM2], there are no assumptions about delays, and the only assumptions about voltage levels deal with the ordering of the sequence in which signals make transitions with respect to standard thresholds; these thresholds can be scaled along with all other voltages. Thus for QDI circuits the same computation is performed regardless of the scaling constants s and α .

Definition 30. For an arbitrary smooth circuit, consider a solution $\mathbf{v}(t)$. Define the *instantaneous dissipative power* $P_{\mathbf{v}}(t)$ as the sum over all transistors of $v_{DS}i_D$. If the circuit halts at time $\tau_{\mathbf{v}}$ on initial condition $\mathbf{v}(0)$, then also define the *energy consumption* as

$$E_{\mathbf{v}} = \int_0^{\tau_{\mathbf{v}}} P_{\mathbf{v}}(t) dt. \quad (31)$$

By Tellegen's theorem [IW], the net power emitted by all capacitors equals the net dissipative power consumed by the transistors. In computing the latter, only source-drain voltages need be considered because there is no gate current. Only the dissipated energy cannot be recovered, so we only consider this energy in our definition of energy consumption.

Energy and Latency Scaling Theorem. Consider an arbitrary smooth circuit for which the solution $\mathbf{v}(t)$ with initial condition $\mathbf{v}(0)$ is unique and halts. Let $E_{\mathbf{v}}$ and $\tau_{\mathbf{v}}$ denote respectively the energy and latency of the computation performed by $\mathbf{v}(t)$. For any positive scalar s , the unique solution $\mathbf{u}(t)$ of the circuit with initial condition $\mathbf{u}(0) = s\mathbf{v}(0)$ halts and performs the same computation, using energy $E_{\mathbf{u}} = s^2 E_{\mathbf{v}}$, and having latency $\tau_{\mathbf{u}} = \tau_{\mathbf{v}}/s$.

Proof: From the voltage scaling law and its corollary, we know that $\mathbf{u}(t) = s\mathbf{v}(st)$. By Assumption 29, \mathbf{u} performs the same computation as \mathbf{v} . Also the latency of \mathbf{u} is

$$\tau_{\mathbf{u}} = \min_{t: \mathbf{u}_{out}(t) > 0} t = \min_{t: s\mathbf{v}_{out}(st) > 0} t = \min_{t: \mathbf{v}_{out}(st) > 0} t = \min_{t: \mathbf{v}_{out}(t) > 0} t/s = \tau_{\mathbf{v}}/s,$$

thus \mathbf{u} halts. The instantaneous dissipative power of \mathbf{u} at time t is:

$$\begin{aligned} P_{\mathbf{u}}(t) &= \sum_{\text{transistors}} v_{DS}(\mathbf{u}(t)) \cdot f_D(\mathbf{u}(t)) \\ &= \sum_{\text{transistors}} v_{DS}(s\mathbf{v}(st)) \cdot f_D(s\mathbf{v}(st)) \\ &= \sum_{\text{transistors}} s v_{DS}(\mathbf{v}(st)) \cdot s^2 f_D(\mathbf{v}(st)) \\ &= s^3 \sum_{\text{transistors}} v_{DS}(\mathbf{v}(st)) \cdot f_D(\mathbf{v}(st)) \\ &= s^3 P_{\mathbf{v}}(st), \end{aligned} \quad (32)$$

so the energy consumed by \mathbf{u} is

$$E_{\mathbf{u}} = \int_0^{\tau_{\mathbf{u}}} P_{\mathbf{u}}(t) dt = \int_0^{\tau_{\mathbf{v}}/s} s^3 P_{\mathbf{v}}(st) dt = s^2 \int_0^{\tau_{\mathbf{v}}} P_{\mathbf{v}}(t) dt = s^2 E_{\mathbf{v}}. \quad \square \quad (33)$$

Constant $E\tau^2$ Corollary. For an arbitrary smooth circuit for which the solution $\mathbf{v}(t)$ with initial condition $\mathbf{v}(0)$ is unique and halts, the computation performed and $E\tau^2$ is independent of scaling of $\mathbf{v}(0)$.

Proof: let $\mathbf{u}(0) = s\mathbf{v}(0)$, and consider the unique solution $\mathbf{u}(t)$. By the Energy and Latency Scaling Theorem, the same computation is performed, $E_{\mathbf{u}} = s^2 E_{\mathbf{v}}$, and $\tau_{\mathbf{u}} = \tau_{\mathbf{v}}/s$. Therefore

$$E_{\mathbf{u}} \tau_{\mathbf{u}}^2 = (s^2 E_{\mathbf{v}}) \cdot (\tau_{\mathbf{v}}/s)^2 = E_{\mathbf{v}} \tau_{\mathbf{v}}^2. \quad \square \quad (34)$$

Since the freedom of choosing s in the definition of \mathbf{u} captures all variations in initial conditions which we are considering for a particular computation, we will drop the subscripts and attribute $E\tau^2$ to the computation itself. More precisely, we attribute $E\tau^2$ to the circuit for a class of inputs which result in a particular computation. Even if τ changes through voltage scaling, $E\tau^2$ remains constant.

Constant $E\tau^2$ is a result of the voltage scaling law and uniqueness of circuit solutions. Therefore the result applies to ordinary smooth circuits. The result also applies to smooth circuits with additional capacitors, provided circuit solutions are known to be unique.

Conclusion

We have shown that circuits exhibiting a voltage scaling law have constant $E\tau^2$. In particular, we have defined the family of “smooth circuits” to which this result applies. In constructing this family, we have demonstrated the simplicity of analysis which results from generically adding a capacitor from each node to ground.

Although we would generally like to use $E\tau^2$ as a performance metric, we ought to point out certain assumptions that were made during the construction of this family. These assumptions ought to be noted when applying the $E\tau^2$ metric to real circuits.

1. We assumed that every node has a capacitor to GND, and that the power supplies can be modeled as large capacitors. Also we assumed that all inputs are encoded in initial voltage conditions. In particular, the input voltages must be scaled along with the supply voltage. And there is no output current for the circuit as a whole. These assumptions ought to be satisfiable in practice.
2. We assumed that circuit solutions are unique. In practice, noise makes it difficult to compare any particular physical circuit solution to a unique desired solution: even if the amount of noise is small and the circuit logic is highly restoring, absolute drifts in time occur routinely; in fact these drifts are acceptable behavior in circuit design styles such as QDI. However, even though the designer cannot guarantee that the physical circuit solution matches a particular solution, the designer has identified a set of *acceptable* solutions. Each element of this solution set can be characterized by E and τ , and the energy and latency scaling theorem can be applied to each solution. If linear statistics (such as a weighted mean) are then used to define E and τ of the entire solution set, then these aggregate E and τ will also scale according to the energy and latency scaling theorem, and $E\tau^2$ will remain constant.
3. We assumed that the threshold voltage can be scaled along with the supply voltage. Circuit designers may not always have control over this, or might prefer higher threshold voltages for noise rejection.
4. We assumed that the circuit was comprised of ideal MOS transistors. Ordinary resistors, and MOS transistors with velocity saturation or the Early effect do not satisfy the square law, so adding these types of components to the circuit will cause $E\tau^2$ to change when V is changed.

Finally, we have suggested that by changing the latency τ , we are changing the speed of the circuit. Besides latency, another common measure of speed is throughput. There are at least two ways of seeing that we are scaling latency and throughput together:

1. The voltage scaling law scales the entire trace temporally. Thus if the throughput is measured as the rate of oscillation of a particular signal in that trace, then the throughput will increase as the latency decreases.
2. One way of measuring throughput is by building a test circuit that processes a large number of messages sequentially, and then dividing this circuit’s measured latency by the number of items. The throughput of this test circuit thus scales together with its latency.

Bibliography

- [AM] A. J. Martin. Towards an Energy Complexity of Computation. *Information Processing Letters*, 77, 2001, pp. 181-187.
- [AM2] A. J. Martin. Synthesis of Asynchronous VLSI Circuits. *Formal Methods for VLSI Design*, ed. J. Staunstrup, North-Holland, 1990.
- [NM] Norbert R. Malik. *Electronic Circuits: Analysis, Simulation, and Design*, Prentice Hall, 1995.
- [HW] J. Hubbard and B. West. *Differential Equations: A Dynamical Systems Approach*, Springer-Verlag New York, 1991
- [IW] Irwin and Wu. *Basic Engineering Circuit Analysis*, Prentice Hall, 6th edition.